

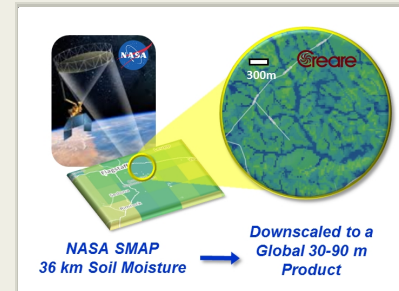
Open-Source Pipeline for Large-Scale Data Processing, Analysis and Collaboration, Phase I

Completed Technology Project (2017 - 2017)



Project Introduction

NASA's observational and modeled data products encompass petabytes of earth science data available for analysis, analytics, and exploitation. Unfortunately, these data are underutilized due to vast computational resource requirements; disparate formats, projections, and resolutions that hinder data fusion and integrated analyses; complex and disjoint data access and retrieval protocols; and task-specific and non-reusable code development processes that hinder algorithm sharing and collaboration. Due to these limitations, recent advances in unsupervised machine learning using deep neural nets (DNNs) have not been widely adopted for applications such as pixel-based classification, image preprocessing, feature recognition, and scene understanding. Creare proposes to develop an open-source, standards-based Python software framework that removes major barriers to widespread exploitation of geospatial earth science data. This will be achieved through development of PODPAC (Pipeline for Observational Data Processing, Analysis, and Collaboration), a pipeline-based architecture that (1) enables multi-scale and multi-windowed access, exploration, and integration of available earth science data sets to support both analysis and analytics; (2) automatically accounts for differences in underlying geospatial data formats, projections, and resolutions; (3) simplifies implementation and parallelization of geospatial data processing routines; (4) seamlessly integrates with DNN machine learning frameworks; and (5) unifies access, processing, and sharing of data and algorithms via interfaces to existing NASA repositories. To demonstrate the impact of these innovations, we will use PODPAC to derive an on-demand, high-resolution global soil moisture data product from the Soil Moisture Active/Passive (SMAP) satellite radiometer observational data to support applications in hydrology, agriculture, and humanitarian response missions involving flooding, drought, and water resources.



Open-Source Pipeline for Large-Scale Data Processing, Analysis and Collaboration, Phase I Briefing Chart Image

Table of Contents

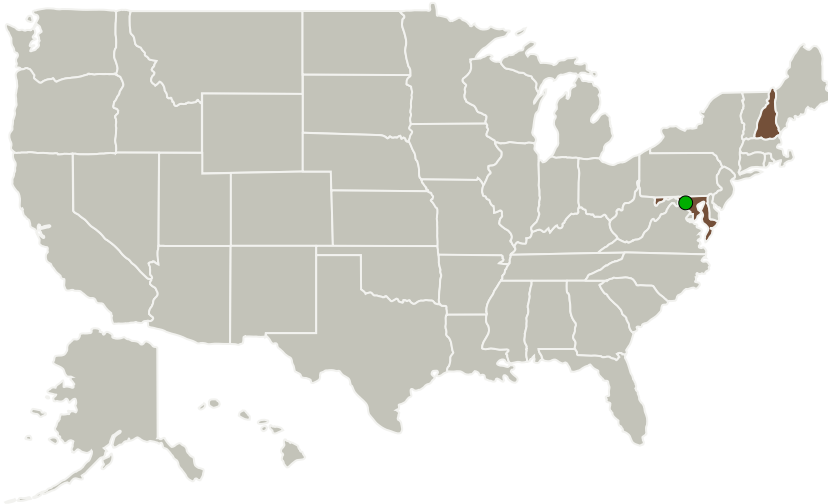
Project Introduction	1
Primary U.S. Work Locations and Key Partners	2
Organizational Responsibility	2
Project Management	2
Technology Maturity (TRL)	2
Images	3
Technology Areas	3


Open-Source Pipeline for Large-Scale Data Processing, Analysis and Collaboration, Phase I

Completed Technology Project (2017 - 2017)



Primary U.S. Work Locations and Key Partners



Organizations Performing Work	Role	Type	Location
Creare LLC	Lead Organization	Industry	Hanover, New Hampshire
 Goddard Space Flight Center(GSFC)	Supporting Organization	NASA Center	Greenbelt, Maryland

Primary U.S. Work Locations

Maryland	New Hampshire
----------	---------------

Organizational Responsibility

Responsible Mission Directorate:

Space Technology Mission Directorate (STMD)

Lead Organization:

Creare LLC

Responsible Program:

Small Business Innovation Research/Small Business Tech Transfer

Project Management

Program Director:

Jason L Kessler

Program Manager:

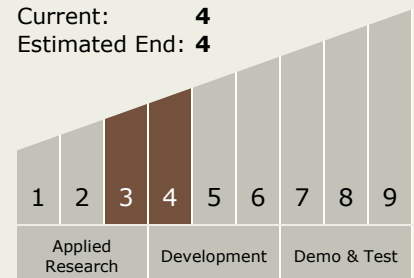
Carlos Torrez

Principal Investigator:

Jerry Bieszczad

Technology Maturity (TRL)

Start: **3**
 Current: **4**
 Estimated End: **4**

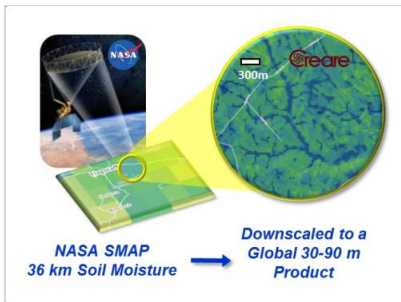


Open-Source Pipeline for Large-Scale Data Processing, Analysis and Collaboration, Phase I

Completed Technology Project (2017 - 2017)



Images



Briefing Chart Image

Open-Source Pipeline for Large-Scale Data Processing, Analysis and Collaboration, Phase I Briefing Chart Image

(<https://techport.nasa.gov/image/130574>)

Technology Areas

Primary:

- TX11 Software, Modeling, Simulation, and Information Processing
 - └ TX11.4 Information Processing
 - └ TX11.4.2 Intelligent Data Understanding